

Title	英語コーパスを用いた語彙使用の変遷に関する実証的研究
Sub Title	A corpus-based study of the changes in the usage of lexical items in English
Author	佐野, 真一郎(Sano, Shinichiro)
Publisher	慶應義塾大学言語文化研究所
Publication year	2022
Jtitle	慶應義塾大学言語文化研究所紀要 (Reports of the Keio Institute of Cultural and Linguistic Studies). No.53 (2022. 3) ,p.59- 83
JaLC DOI	
Abstract	This paper presents a corpus-based study examining the recent changes in the English lexicon due to the development of technological advancements and globalization from the mid-twentieth century onwards. Specifically, focusing on eight lexical items, this study investigates the chronological transition and semantic changes in the use of these items as well as differences in style and register. The results confirm the influence of extra-linguistic factors on the recent changes in the use of lexical items: 1) all items show a rapid increase in frequency from around the 1990s; 2) the technological meaning of “mouse” and “web” has spread from the 1990 s; 3) although “internet” and “web” are synonymous, there is a stylistic difference: “internet” is more compatible with formal contexts; 4) “web” is more frequent in the informational register than in others.
Notes	論文
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00069467-00000053-0059

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

英語コーパスを用いた語彙使用の変遷に関する実証的研究¹

佐野 真一郎

Abstract

This paper presents a corpus-based study examining the recent changes in the English lexicon due to the development of technological advancements and globalization from the mid-twentieth century onwards. Specifically, focusing on eight lexical items, this study investigates the chronological transition and semantic changes in the use of these items as well as differences in style and register. The results confirm the influence of extra-linguistic factors on the recent changes in the use of lexical items: 1) all items show a rapid increase in frequency from around the 1990s; 2) the technological meaning of “mouse” and “web” has spread from the 1990s; 3) although “internet” and “web” are synonymous, there is a stylistic difference: “internet” is more compatible with formal contexts; 4) “web” is more frequent in the informational register than in others.

1. はじめに

これまで、英文学、英語学、語源学、辞書学やその周辺領域における取り組みによって、英語の語彙・音韻・形態・統語的特徴とそれらの変遷が明らかにされてきた（Barber et al. 2009, Baugh and Cable 2002, Crystal 1995, McArthur 1992）。

1 本稿は、佐野（2020）の一部の内容に基づき、加筆修正したものである。

言語の語彙は、様々な言語的・非言語的要因によって形作られる。その中で、言語接触に伴う借用や接辞による派生、複合語形成などにより新たな語彙が追加される (Hughes 2008, Trask 1996, Traugott 1982)。英語に目を向けると、古英語ではゲルマン語起源の語彙が支配的であり、借用語は少なく、複合語が多かった。一方、中英語では言語接触によるフランス語やラテン語からの借用語が語彙拡張の主な要因となった。また、借用語の中でも、接触の形態・特徴によって取り込まれるレジスターが異なる。これらの結果、英語の語彙は主に日常語・一般語彙を担うAnglo-Saxon (本来語)、法律・行政に関する語彙を担うNorman-French、宗教や学術・専門用語を担うLatin/Greekの3つの層から成る (Hughes 2008)。

語彙全体の増減だけでなく、各語彙項目の意味・用法の変化も生じる。これまでの研究において、語彙の意味変化のパターンが特定され、類型化が行われてきた (Arlotto 1972, Bloomfield 1933, Crowley 1992, Geeraerts 1997, Stern 1931)。例えば、意味の良化と悪化、拡張・一般化と縮小・特殊化などが代表的なものとして挙げられる。また、意味変化の駆動力として、隠喩や換喩が関わっており、伝統的な意味が比喩的に転用されることによって新たな概念を表すに至る (Riemer 2010, Traugott 1989)。

このような語彙の変化には、歴史・文化・社会・心理的要因や、言語使用における単語同士の相互作用など様々な要因が関わっている (Sheard 1966, Traugott 1982, Trask 1996, Ullmann 1962)。その中でも、19・20世紀以降の語彙変化に多大な影響を与えたのは技術革新である。産業革命は人の移動を引き起こし、方言同士の接触へと繋がった。それ以降も、郵便、電報、電話、ラジオ、テレビに代表されるコミュニケーション技術の発展により、言語・方言接触が促進された。更に、情報技術革命はインターネットやSNSを主要なコミュニケーション手段にし、同時に国際化も加速した。このような技術の発展によって、レジスターも多様化し、新たな概念の受け皿となる新規語彙が生み出されてきている (例、HTML (Hyper Text Markup Language) などの頭字語や省略語)。またこの変化は、既存語彙の意味・用法の変化も伴う (例、flaming: インターネット上の炎上) (Mincova and Stockwell 2006)。ラ

テン語・ギリシャ語を語根とする専門用語（例, computer）も増加している。この結果, 英語の語彙構成に変化が生じ, 今や使用頻度の高い上位600語の多くは, 組織・ビジネス関連語彙（例, company, manager）, 専門用語（例, energy, machine）であり, かつて中心的であった宗教・道徳の語彙（例, honour, virtue）は全く含まれていない（Hughes 2008）。

近年のこうした語彙変化に関する具体的な事例研究として, Andersson (2020) は, web, cloud, mouse, bugを取り上げ, コーパスを用いたコロケーションと文脈の分析結果（1950年代から1980年代までは従来の意味・用法が支配的であるが, 1990年代ごろから情報技術関係の意味・用法が増加している）を一般化し, 語彙による変化の進み方の違いや例外はあるものの, 情報技術に関する意味が, 従来の意味に追加され（意味拡張, semantic widening）, 多義性を持つ状態（polysemy）を経て, 従来の意味に取って代わる（意味の交代, replacive change）という, 技術革新による語彙の意味変化の方向性を予測している。

しかしながら, 英語を含めた語彙とその変化の研究ではこれまで類型化に主眼が置かれ, 近年の技術革新による語彙変化はもとより, 実際の言語使用に基づく事例研究・実態調査には十分な蓄積が無い。以上を踏まえ, 本研究ではコーパスを用いて 1) 19世紀から現在までの単語使用の実態, 2) 意味・用法の変化, 3) スタイルによる違い, 4) レジスターによる違いについて実証することを目的とする。

2. 方法

2.1 調査対象

現代の国際化の進展や情報技術の発展により, 特定語彙の使用頻度が増加する, 新たな用法を獲得する, または語彙が新規に生み出されるなどのことが予想される。本研究では, このような変化が予想される以下の語彙8種を対象とした。

名 詞： globalization, internet, blog, computer, mouse, web

形容詞： international, online

この中で、internationalとglobalizationは国際化と関係が深いと考えられる。internet, blog, computer, mouse, web, onlineは、情報技術と関係の深い語彙に分類できる。これらの語彙の選定において、専門用語を対象とした場合、分析に耐えるだけの用例数が得られるのは、特定のレジスターを対象とする「特定目的コーパス」に限られる可能性が高く、英語の語彙使用実態の記述としては一般性が低下してしまう恐れがある。従って、より一般的な「汎用コーパス」でも一定の用例数が確保できるよう、専門用語ではなく、一般に普及している単語を選んだ。

また、数ある英語変種・方言の中でも、アメリカ英語、およびインターネット上の英語を対象とすることとなった。これは、調査目的に合致するコーパスの対象が、それぞれアメリカ英語・インターネット上の英語であったという実用上の理由によるものである。

2.2 コーパス

本研究では、分析目的に応じて3種類の英語コーパスを併用した。ここで各コーパスの特徴を紹介する。

COHA (Corpus of Historical American English) は、1820年から2019年までのアメリカ英語を収録したコーパスである²。4億7,500万語規模のテキストを収録しており、構成比率としてFictionに偏りがあるものの(塚本 2011), Fiction, Popular Magazines, Newspapers, Non-Fiction Booksの4レジスターから収集されており、後期近代から現代までのアメリカ英語の使用実態を表した、代表性を備えるコーパスと概ね見ることができる。本研究では、COHA

2 English-Corpora.org上で、COHAが収録しているデータの対象期間は、2020年8月の段階では1810年代から2000年代(～2009年)までであったが、2021年8月の段階では1820年代から2010年代(～2019年)までとなっていた。本研究では、1820年代から2010年代までを対象とした。

を用いて語彙8種の使用実態の通時的変遷を記述するとともに、一部の単語についてコロケーションの分布から意味・用法の変化に関する証左を得た。

Corpus of US Supreme Court Opinionsは、1790年代から現在までのアメリカ英語を収録した1億3,000万語規模のコーパスである。本コーパスの特徴として、アメリカ連邦最高裁の判例集に含まれる文書を対象としており、法律文書という特定のレジスターを代表する言語使用データを提供する点が挙げられる。本研究では、一部の単語についてCorpus of US Supreme Court Opinionsにおける分布とCOHAにおけるそれを比較し、単語使用のスタイル差について検証した。

Corpus of Online Registers of Englishは、2014年のインターネット上の英語5,000万語を収録している。これらの言語使用データは、Oral, Blog, Narrativeなどを大分類として、30種類以上のレジスターに分類されており、異なるレジスター間の分布特徴を比較できる。本研究では、Corpus of Online Registers of Englishを用いて、一部の単語についてレジスター差の比較を行った。

2.3 データ検索

本研究では、データ検索インターフェースとして、English-Corpora.org (<https://www.english-corpora.org/corpora.asp>)を用いた。English-Corpora.orgは、英語コーパスを対象とするオンライン検索システムであり、各コーパスから目的とする表現をウェブ上で検索することができる。その結果を集計・表示する簡易的機能も備える。同システムでは、現在BNC, COCA, COHAを始めとする17種類の英語コーパスが利用可能である。これらのコーパスは、それぞれ性質が異なっており、話し言葉を対象とするものもあれば、書き言葉を対象とするものもある。その他、対象とする時代、レジスター、地域・方言についても様々である。例えば、1470年代から1690年代の英語を対象とするEarly English Books OnlineやCOHAなどの歴史コーパスを用いることで、通時変化などの調査ができる。レジスターについては、BNCはイギリス英語の話し言葉・書き言葉（フィクション、雑誌、新聞）を収録しているが、Corpus of Online Registers of Englishは多様なレジスターにわたるウェブ上の

記事を、Corpus of US Supreme Court Opinionsは法律文書を対象としており、レジスターごとの特徴の比較などができる。Global Web-Based Englishは、20カ国の英語を収録しており、英語変種・方言の違い、地域差などを比較できる。その他、100億語を超える規模の「超大規模コーパス」であるNews on the Webも利用することができる。同コーパスは、2010年から現在までのウェブ上の新聞・雑誌記事を収録している。その規模は131億語（2021年8月5日現在）であるが、データを更新し続けるモニターコーパスであるため、毎年約20億語ずつ増加している。

このように、異なる複数のコーパスを同一のプラットフォームで検索することで、コーパス間比較が比較的容易に行えるだけでなく、研究対象について同一基準で多角的に分析・検証ができる。

English-Corpora.orgでは、文字列検索に加えて、品詞・lemma検索、コーパスのどのセクションを対象とするかなどの指定をすることができるが、今回は文字列検索以外の指定はせずデフォルトのままとした。その他、同システムは検索に関わる以下の機能を備えている。

- ・ List：検索語の頻度を表示する。
- ・ Chart：検索結果を項目別にグラフ表示する。
- ・ Collocates：単語同士の共起例（コロケーション）を表示する。
- ・ Compare：2種類の単語の共起例を並列表示する。
- ・ KWIC (keyword in context)：検索語の前後文脈を併せて表示する。

本研究では、分析目的別にChartとCollocatesを使用した。詳細については第3節で述べる。

3. 分析

前節の方法に従い、各コーパスから必要な言語使用データを収集した。

データは、2021年8月4日から5日にかけて取得した³。収集したデータは、下記4項目について分析した。

1. 通時的変遷 (COHA)
2. 意味・用法の変化 (COHA)
3. スタイルによる違い (COHA, Corpus of US Supreme Court Opinions)
4. レジスターによる違い (Corpus of Online Registers of English)

以下、項目別に分析結果の説明と考察を行う。

3.1 通時的変遷

まず、Corpus of Historical American English (COHA) を用いて、1820年代から2010年代までのアメリカ英語における単語使用の変遷を分析する。ここでは、検索結果を項目別にグラフ表示するChart機能を用いた。検索の際にSectionsにチェックを入れることで、結果をSection（ここでは年代）ごとに分類して表示した。Chart機能が出力するのは、1) 検索語の頻度、2) コーパスの総語数、3) 100万語あたりの調整頻度、4) 調整頻度の棒グラフである。調整頻度は、検索語の頻度をコーパスの総語数で割ったものである。検索語の頻度の価値は、コーパスの総語数によって変わるため（例えば、同じ1件でも1万語あたりの1件と10万語あたりの1件では意味が異なる）、本研究では、主に調整頻度に基づき議論を進める。

表1は、各単語の頻度（調整頻度）を年代ごとにまとめたものである。なお、煩雑さを避けるため、English-Corpora.orgから直接得られた結果は、付録にまとめた。

3 English-Corpora.orgに含まれるコーパスには、常にデータが更新されているものもある。従って、データ取得時期により表示される検索結果が異なる可能性がある。

表1. COHAにおける語彙頻度の通時的変遷（数値は100万語あたりの調整頻度）

	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910
international	2.44	4.96	5.63	7.32	20.72	15.33	18.99	26.19	37.49	63.19
globalization	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
computer	0.00	0.00	0.00	0.00	0.00	0.16	0.05	0.00	0.09	0.04
internet	0.00	0.00	0.06	0.36	0.00	0.11	0.10	0.24	0.09	0.00
online	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
blog	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mouse	1.72	2.77	2.85	8.41	9.03	8.84	7.67	8.96	9.01	14.76
web	5.30	6.27	6.52	4.72	9.09	7.93	6.08	5.04	5.64	5.06

	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
international	103.42	100.80	114.53	109.49	101.26	95.32	121.53	108.66	116.39	122.33
globalization	0.00	0.00	0.00	0.00	0.00	0.00	0.23	2.65	7.81	4.74
computer	0.04	0.14	0.33	3.42	22.29	51.16	106.39	134.90	105.22	87.52
internet	0.19	0.07	0.04	0.03	0.00	0.10	0.07	40.03	77.05	57.32
online	0.00	0.04	0.07	0.21	0.03	0.14	0.50	21.27	57.35	97.79
blog	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.06	6.06	15.91
mouse	16.54	9.42	13.47	8.62	9.07	10.72	14.20	20.18	21.37	19.27
web	4.79	4.76	4.45	5.41	4.88	5.79	8.48	36.86	72.31	39.35

表1の分布特徴について、網掛け部分を中心に見てゆく。まず、国際化と関係が深いと考えられるinternationalとglobalizationの分布についてまとめる。internationalは、1900年代ごろから増加し、1920年代から現在まで100ポイント以上の数値を保っている。一方、globalizationは1980年代から観察され、そこから大幅に増加しているものの、他の語彙と比べるとその頻度は低い⁴。また、両語彙を比較した場合、internationalは全年代を通して用例が観察されることから、調査語彙8種の中では長い歴史を持つのに対し、globalizationは近年普及し始めた語彙であると言える。Online Etymology Dictionary (<https://www.etymonline.com/>)によると、internationalは1780年が初出とされる造語であり、200年以上の歴史がある。globalizationのglobalからglobalizeを経て派生した道筋を追うと、globalは1892年が初出であり、比較的早くから存在するが、globalizeは1950年代、globalizationは1960年代から使われ始めている。このことから、「世界的な」、「地球規模の」という概念は長らく存在

4 2010年代で、globalization, computer, internet, webの頻度が低下しているが、これに関する明確な理由は分からなかった。

するが、その変化を含むグローバル化という概念自体が近年生じた可能性が示唆される。また、**globalize**, **globalization**という派生が生じるには、語基となる**global**の使用頻度がある程度高まらなければならないという可能性も考えられる。実際、COHAで**global**を検索すると、1920年代から1970年代まで0.27～9.75ポイントへとわずかに増加しているが、大幅に増加するのは1980年代以降である（1980年代：18.89, 2010年代：80.87）。この**global**の大幅な増加と連動するように、**globalization**の頻度も上昇しているのである。いずれにしても、**international**と**globalization**の分布特徴は、近年国際化が加速している、または少なくとも国際化についての話題がメディア等で取り上げられる機会が増加していることを反映していると考えられる。

次に情報技術と関係が深いと考えられる語彙の分布をまとめる。**computer**は1960年代から急激に増加し、1980年代に100ポイントを超え、それ以降高い頻度を保っている。Online Etymology Dictionaryによると、計算機ではなく、今日的なコンピュータの意味での**computer**の初出は1945年となっており、それ以降COHAにおける使用頻度も増加していることが分かる。一方、情報技術の中でもインターネットと関係する**internet**, **online**, **blog**は、1990年代、2000年代から増加している⁵。また、これらは似通った分布を示しているが、頻度を比べた場合、**blog**が最も低い値（1.06～15.91ポイント）を示していることから、この中では最も新しい語彙であると考えられる。実際、Online Etymology Dictionaryでは1998年が初出となっている。これは、**blog**が**web**と**log**を基にする混成語であり、同じくインターネット関連の**web**の普及が**blog**成立の要件となることと、SNSとしての「ブログ」が普及するのはインターネット自体の普及より後であることから説明できるだろう。ここまできをまとめると、情報技術に関係のある語彙の中でも、**computer**は比較的早く、20世紀後半から使われ始め、増加している。一方、インターネット関連

5 COHAにおける**internet**の1800年代の用例は、Internet Archiveのようなものばかりであった。これらは1800年代の資料を収録する（近年）編纂されたアーカイブの注釈内の用例である。従って、**internet**が1800年代に使用されていたということではない。

のinternet, online, blogは、その技術が一般家庭に広まった1990年代以降に普及したと言える。

最後にmouseとwebの分布を見ると、他の情報技術関連の用語と同様1990年代以降頻度が高くなっている。中でもwebは、1980年代以前は10ポイント未満であるのに対して、1990年代以降は36ポイント以上となっており、使用頻度の顕著な増加が見られる。しかしながら、これらは同時に1820年代から一定数の用例が比較的安定して観察される。mouseについては、語源的に古英語mus, 更にはインドヨーロッパ祖語**mus*にまで遡れるように、それ自体長い歴史のある語彙であるが、現代的なコンピュータとその周辺機器である「マウス」が登場したのは20世紀である。従って、20世紀前半以前のmouseの用例は情報技術とは異なる用法である可能性が高い。また、webも長い歴史を持つ（古英語webb, インドヨーロッパ祖語*(h)uebh-）。情報技術としてのwebは、World Wide Webの省略語として1990年代に使われ始めたことから、1980年代以前の用例は情報技術とは異なる用法と考えられる。このmouseとwebの用法の変化（意味変化）の可能性について、次節ではコロケーションを手がかりに検証する。

3.2 意味・用法の変化

mouseとwebは、元々中心的な意味として「ねずみ」「蜘蛛の巣・糸」を意味するが、同時に（パソコンの周辺機器としての）「マウス」「インターネット」の意味も持つ。そうすると、表す意味・用法によってそれぞれ共起する単語が異なるはずである。ここでは、コロケーションの分布を年代別に見ることによって、意味変化の様相を視覚化する。情報技術関連の意味（マウス、インターネット）であれば1990年代以降、つまり前節で頻度の増加が見られた時期に分布し、伝統的な意味（ねずみ、蜘蛛の巣・糸）であれば1820年代から2010年代まで広く分布するという仮説を立てる。

検索には、COHAのCollocates機能を用いた。Word/phraseの欄に検索語を、Collocatesの欄にはワイルドカード（制限を設けない）の「*」を入力し、結果を年代別に分類表示するためSectionsにチェックを付けた。その他の項目

表2. COHAにおけるmouseの共起語の通時的変遷

	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	ALL	%	MI		
MICKEY	358												35	16	15	32	61	58	63	49	29	3207	11.16	10.15		
CAT	238			2	11	4	8	13	16	13	12	19	17	12	17	10	13	14	12	30	15	19935	1.19	6.93		
MOUSE	172				4		6		24	8	4	2	8	18	10	2	4	16	22	30	14	5840	2.95	8.23		
SINGING	120										118											23150	0.52	5.72		
MEADOW	112				2	1	2	1	1	1	3	101					1				1	5059	2.21	7.82		
RAT	66				9	2	1	3	4	3		2	1		15	4	3	1	6	7	5	5748	1.15	6.87		
QUIET	63			2	1	3	4	4	3	4	7	2	2	3	7	4	2	3	2	7	5	53988	0.12	3.57		
TRAP	56		2		1	1	1	5	5	1	7	10	4	4	4	2	1	1	5	3	3	8348	0.67	6.10		
FIELD	54					3	2	1	2	1	4	3	3	6	5	4	3	1	5	8	3	67883	0.08	3.02		
GRAY	53									3	1	31		1		6	1	2	1	1	6	41243	0.13	3.71		
TINY	49				1	6	1	1	1		6	2	1	1	4	2	2	2	3	8	9	25999	0.19	4.26		
HOLE	40			1		2	1	6		1	2	1	2	2	2		4	5	4	4	3	22200	0.18	4.20		
CLICK	39																	1	14	15	9	3998	0.98	6.64		
LION	36				4	3	3	2	1	6	6	2	2	2	1	1	2	1			2	9493	0.38	5.27		
KEYBOARD	33												1									1357	2.43	7.95		
TAIL	30			1	1	1	1	3	4			5		2		1	2	2	2	5	2	14177	0.21	4.43		
EARS	29					1		1	1	1		1			1	2	1	4	9	4	4	29959	0.10	3.30		
CELLS	28													2	1	1		1	2	8	13	12989	0.22	4.46		
COMPUTER	28																	2	7	10	9	16665	0.17	4.10		
MINNIE	26													2	1	2	2	3	7	6	3	1951	1.33	7.09		
MODEL	26											1									1	1	23	21758	0.12	3.61

はデフォルトとした。

本節では、COHAのコロケーション検索で得られた結果のうち、頻度の高かった共起語上位20語の年代分布をそのまま示す。まず、表中の情報について整理する。左列から、a) 共起語のリスト、b) 共起語の頻度、c) 共起語の各年代における頻度、d) 共起語単独での頻度、e) d)に対するb)の比率（ $(b \div d) \times 100$ ）、f) MIスコアとなる。MI (mutual information, 相互情報量) スコアは、コロケーションの結び付きの強さを示す統計的指標の一つであり、対象となる語(検索語)が共起する語の情報をどの程度持っているかを示す。数値が大きいほど強い結び付きを意味する(石川 2006)。Hunston (2002) に倣い、本研究でもMIスコアが3以上であるものをコロケーションと見做すという基準を採った。

mouseのコロケーションについてまとめた表2を見ると、mickey, cat, ratなどと共起する「ねずみ」の用例は1820年代から幅広く分布しているが、click, keyboard, computerなどと共起する「マウス」の用例が多く観察されるようになるのは1990年代以降に限られる。つまり、仮説の通り伝統的な意味(ねずみ)が1820年代から現在まで使用される中で、情報技術関連の意味(マウス)が新たに加わり、1990年代以降普及してきていると考えられる。Andersson (2020) の調査でもmouseの意味拡張が確認されており、本研究でも同様の結果となった⁶。ただし、全体の分布を見ると、20語のうち依然として伝統的な「ねずみ」を指す共起語がほとんどであることから、情報技術関連の意味は現時点では様々な語彙と共起するまでには普及しておらず、一部の語彙に偏っていると言える。

次にwebのコロケーションに関する表3を見ると、spider, spun, tangled, weaveなどと共起する伝統的な「蜘蛛の巣・糸」の用例は1820年代から幅広く見られる。一方、site(s), page(s), browserなどと共起する「インターネット」の用例はほぼ1990年代以降に限られる。この結果から、仮説の通り伝統的な

6 Andersson (2020) も本研究と同じくCOHAのコロケーションを分析しているが、1950年代と1990年代など二期間の比較であることと、共起語のMIスコアを分析の拠り所としている点で、本研究の方法とは異なる。

表3. COHAにおけるwebの共起語の通時的変遷

	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	ALL	%	MI
SITE																		271	908	179	15890	8.55	9.44
SITES										1								102	258	42	5933	6.79	9.11
SPIDER	3	7	13	14	15	29	16	8	21	10	17	19	18	19	13	17	16	34	30	35	3087	11.47	9.86
*	238	2	1	5	5	1	1	1		2	3		1		2	1		42	175	4	198400	0.12	3.29
WIDE	172														2			101	47	20	47314	0.36	4.89
PAGE	127																	36	57	34	36897	0.34	4.81
TANGLED	94		5	2	1	5	3	6	1	8	3	2	1	5	5	3	14	7	14	8	4313	2.18	7.47
VISIT	91																	27	60	4	44940	0.20	4.04
SPUN	89	1	2	3	3	3	3	7	6	4	5	6	9	7	2	10	6	5	2	5	4384	2.03	7.37
HTTP	76																	46	25	5	1888	4.03	8.35
INFORMATION	74																	2	25	38	56008	0.13	3.43
WOVEN	73	3	5	5	2	8	2	6	6	4	2	7	1	5	3	1	4	2	1	4	3727	1.96	7.32
PAGES	72																	21	37	14	19939	0.36	4.88
WEAVE	70	2	6	2	2	4	5	6	2	2	4	4	2	2	4	3	1	9	2	6	2202	3.18	8.01
CAUGHT	62									1	3	3	3	6	10	9	7	11	5	4	57253	0.11	3.14
E-MAIL	54																	12	38	4	3286	1.64	7.06
BROWSER	51																	25	9	17	365	13.97	10.15
ACCESS	51																	10	28	13	14169	0.36	4.87
WEAVING	47	1	1	3	4	4	6	5	6	1	2	4	4	4	3	1	1	1	1	1	2933	1.60	7.03
INTERNET	47																	18	25	4	6079	0.77	5.97

意味（蜘蛛の巣・糸）が1820年代から現在まで使用される中で、情報技術関連の意味（インターネット）が現れ、1990年代以降普及してきていると考えられる。

mouseの分布とwebの分布を比べると、前者では情報技術関連の意味は限定的であったが、後者では20語のうち半数を占める。これには、mouseがコンピュータの特定の周辺機器のみを指すのに対し、webはインターネットに関わる文脈で幅広く使われるということも影響しているだろうが、変化の段階自体の違いもありそうである。Andersson (2020) は、mouseでは伝統的な意味に新規の意味が加わる意味拡張が起こっているが、webの場合は新規の意味が伝統的な意味に取って代わる意味の交代が起こっていると述べている。これが正しければ、webの伝統的な意味と結び付く共起語彙が頻度および種類の点で減少し、反対に新規の意味と結び付く共起語彙が拡大してきていることが本研究でも観察されたと言える。

3.3 スタイル差の比較

internetとwebは、厳密には異なる情報技術を指すが、Andersson (2020) が、webの「複雑な」「入り組んだ」という意味がインターネット（の複雑さ）に隠喩的に転用され、internetとwebは類義語として使用されるに至ったと述べているように、一般的な言語使用においてはどちらもインターネットを意味する語彙である。ここでは、このようにほぼ同義で使用されるinternetとwebに、用法上の違い、とりわけスタイルに関する違いがあるかどうかについて検証する。

この目的のために、COHAとCorpus of US Supreme Court Opinions を用いた。両コーパスは対象とする年代の点でほぼ一致しているが、スタイルの点で異なっている。COHAは複数のレジスターから構成され、アメリカ英語の一般的なスタイルを代表していると考えられる。一方、Corpus of US Supreme Court Opinionsは法律文書という特定のレジスターを対象とするため、COHAと比べてよりあらたまったスタイルを代表していると考えられる。従って、両コーパスにおける分布の違いを比較することによって、語彙

表4. COHAにおけるinternetとwebの語彙頻度の通時的変遷
(数値は100万語あたりの調整頻度)

	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910
internet	0.00	0.00	0.06	0.36	0.00	0.11	0.10	0.24	0.09	0.00
web	5.30	6.27	6.52	4.72	9.09	7.93	6.08	5.04	5.64	5.06
	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
internet	0.19	0.07	0.04	0.03	0.00	0.10	0.07	40.03	77.05	57.32
web	4.79	4.76	4.45	5.41	4.88	5.79	8.48	36.86	72.31	39.35

表5. Corpus of US Supreme Court Opinionsにおけるinternetとwebの
語彙頻度の通時的変遷 (数値は100万語あたりの調整頻度)

	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910
internet	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
web	0.62	0.00	0.00	0.30	0.00	0.33	7.70	1.09	0.83	0.00
	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
internet	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.49	86.18	39.42
web	1.10	6.66	3.15	1.71	2.22	1.42	1.81	4.80	28.77	16.78

使用のスタイル差について推し量ることができる。まず、表4にCOHAにおけるinternetとwebの年代ごとの調整頻度の推移を再掲する。

ここでは、インターネット関連の意味が多く使われ始めた1990年代以降の分布（網掛け部分）に注目する。1990年代、2000年代、2010年代のinternetとwebの頻度を比べると、その差は3.17、4.74、17.97ポイントとなっており、2010年代で広がっているものの全体として差は小さい。

続いて表5のCorpus of US Supreme Court Opinionsにおける分布を見ると、4.69、57.41、22.64ポイントとなっている。頻度の低い1990年代では差が小さいものの、全体としてはinternetの方がwebよりも2倍以上高い頻度を示している。

一般的なスタイルを代表するCOHAでは両語彙の差はあまり無かったが、あらたまったスタイルを代表するCorpus of US Supreme Court Opinionsではinternetの方が多く使われているという結果から、あらたまったスタイルではwebよりもinternetの方が使われやすいと考えられる。つまり、internetとwebはスタイルの点で相互補完的に使用されると言える。この理由について

て、それぞれの出自を考えると、internetは、inter-networkを基に1984年から使われるようになった。一方、webはWorld Wide Webの省略形として1990年代から使われるようになった。また、webの「複雑さ」を表す意味がインターネットの意味に隠喩的に転用された（Andersson 2020）ことを踏まえると、まずinternetがインターネットを表す正式な語彙として成立し、その後webがこれとほぼ同じ意味を表す語彙として派生したと考えられる。このことが、語彙使用のスタイル差に反映している可能性がある。

3.4 レジスター差の比較

最後に、webを例として、語彙使用のレジスター差について検証する。検証には、Corpus of Online Registers of Englishを用いた。本コーパスは、様々なレジスターのインターネット記事を対象としており、異なるレジスター間の分布特徴の比較に適している。

検索には、Chart機能を用いた。検索語の欄にwebを入力し、結果をレジスターに分類表示するためSectionsにチェックを付けた。その他の項目はデフォルトとした。表6に検索結果をvertical chart表示したものを示す。まず、表中の情報について整理する。左列から、a) SECTION = レジスターのリスト、b) FREQ = 検索語の各レジスターにおける頻度、c) SIZE(M) = 各レジスターの総語数（単位：100万）、d) PER MIL = 100万語あたりの調整頻度、e) 調整頻度の棒グラフとなる。なお、ここでは全レジスターのうちoral, blog, narrative, informationalに的を絞って表示する。

まず、表6のレジスターの中で、oral, blog, narrativeでは頻度が低く、informationalで高い。その中でも、FAQ information (487.13ポイント)、info description (308.9ポイント)、How-to/instruct (524.11ポイント) がとりわけ高い数値を示している。また、頻度の比較的低かったblogの中でもinformational blogは例外的に高い数値を示している (279.04ポイント)⁷。この

7 表6から割愛したopinionはoral, blog, narrativeよりやや高い値を示していた。hybridには、information, info descriptionを含むものもあり、高い値を示していたが、全体としてレジスターが混在しており、その影響を特定する妨げとなるため検証の

表6. Corpus of Online Registers of Englishにおけるwebのレジスターごとの分布

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
"ORAL"				
Lyrical	10	0.3	39.73	
TV/Movie Script	0	0.0	0.00	
Interview	39	0.5	86.36	
Formal Speech	6	0.1	74.90	
Spoken	11	0.2	48.95	
Interactive Discuss	319	2.8	115.08	
BLOG				
Personal Blog	290	2.9	100.11	
Travel Blog	18	0.3	54.39	
Opinion Blog	1,271	9.7	130.95	
Relig Blog/Sermon	39	1.2	31.31	
Informational Blog	531	1.9	279.04	
NARRATIVE				
News Report/Blog	1,190	8.5	140.14	
Sports Report	76	2.4	32.33	
Short Story	19	0.6	29.98	
Narrative	20	0.4	47.10	
Historical Article	26	0.8	31.38	
INFORMATIONAL				
FAQ Information	290	0.6	487.13	
Descrip Person	110	0.8	138.89	
Encyclopedia Article	133	1.3	101.89	
Research Article	287	1.7	169.13	
Info Description	1,175	3.8	308.90	
How-to/Instruct	734	1.4	524.11	
Recipe	3	0.1	37.38	

結果から、webは、情報提供・技術指南に関する記事でよく使われることが分かった。それぞれの記事の投稿時期は特定できないが、本コーパスが2014年に編纂されたことを踏まえると、webの意味・用法は主にインターネット関連であると予測される。実際、文脈を確認したところ、ほとんどがインターネット関連の用法であった。従って、インターネットの記事に関する限り、webはインターネット関連の情報提供・技術指南の文脈で多く使われる傾向があることが分かった。

4. 結語

本研究では、現代の国際化の進展と情報技術の発展の影響を受け、近年使用が拡大していると思われる英語の語彙8種を取り上げ、単語使用の通時的変遷、意味・用法の変化、スタイル差、レジスター差について複数のコーパスを用いた数量的分析を行った。その結果、1) 全ての語彙について1990年代以降使用頻度が増加している、とりわけglobalization, internet, online, blogは1990年代ごろから使用が一般化した、2) mouse, webの「マウス」「インターネット」といった情報技術関連の意味は、1990年代ごろから使われるようになった、3) internetとwebはスタイルの点で相互補完的に使用され、あらたまったスタイルではwebよりもinternetの方が使われやすい、4) webはインターネット関連の情報提供・技術指南の文脈で多く使われる傾向があることが分かった。このように、国際化の進展や技術革新といった非言語的要因が、近年の語彙使用の変化に影響を与えていることを確認した。

オンライン検索システムを利用し、性質の異なる複数のコーパスを比較することで、分析対象の特徴を多角的・効率的に分析することができるなど、コーパスのオンライン利用は言語研究の可能性を拡大すると言える。

今後の課題として、本研究では調査対象とした語彙が8種類と少なかったため、調査の範囲を広げる必要がある。また語彙の選定の際には、恣意的で

対象から除いた。

はない、より明確な基準を設ける必要があるだろう。

<謝辞>

本研究は、国立国語研究所共同研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」、科学研究費#19K00558, #16H03426の助成を受けて行った研究成果の一部である。

参考文献

- Andersson, Johan. (2020) Bug report: A study of semantic change during the digital revolution. *Bachelor Degree Project*. University of Stockholm.
- Arlotto, Anthony. (1972) *Introduction to historical linguistics*. Boston: Houghton Mifflin.
- Barber, Charles, Joan C. Beal, and Philip A. Shaw. (2009) *The English language: A historical introduction*, 2nd edition. Cambridge, UK: Cambridge University Press.
- Baugh, Albert C., and Thomas Cable. (2002) *A history of the English language*, 5th edition. London: Routledge.
- Bloomfield, Leonard. (1933) *Language*. New York: Henry Holt and Company.
- Crowley, Terry. (1992) *An Introduction to historical linguistics*, 3rd edition. Auckland: Oxford University Press.
- Crystal, David. (1995) *The Cambridge encyclopedia of the English language*. Cambridge, UK: Cambridge University Press.
- Geeraerts, Dirk. (1997) *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford: Clarendon Press.
- Hughes, Geoffrey. (2008) A history of the English lexicon. In Haruko Momma and Michael Matto (eds.), *A companion to the history of the English language*. Chichester, UK; Malden, MA: Wiley-Blackwell.
- Hunston, Susan. (2002) *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.
- 石川慎一郎 (2006) 「言語コーパスからのコロケーション検出の手法－基礎的統計値について－」『統計数理研究所共同研究レポート』190, 1-14.
- McArthur, Tom. (1992) *The Oxford companion to the English language*. Oxford; New York: Oxford University Press.
- Mincova, Donca and Robert Stockwell. (2006) English words. In Bas Aarts and April McMahon (eds.), *The handbook of English linguistics*, 461-482, Oxford: Blackwell.
- Riemer, Nick. (2010) *Introducing semantics*. Cambridge University Press.
- 佐野真一郎 (2020) 「コーパスのオンライン利用と言語研究の可能性」『日本英文学会第74回中部支部大会プロシーディングス』 <http://www.elsj.org/chubu/2020_proceed>

ings/sano_shinichiro.pdf>

- Stern, Gustaf. (1931) *Meaning and change of meaning*. Bloomington: Indiana University Press.
- Sheard, John A. (1966) *The words of English*. New York: Norton.
- Trask, Robert L. (1996) *Historical linguistics*. London: Arnold.
- Traugott, Elizabeth. (1982) *Some semantic-pragmatic aspects of grammaticalization*. Amsterdam: John Benjamins.
- Traugott, Elizabeth. (1989) On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, 65(1), 31-55.
- 塚本聡 (2011) 「2つの史的コーパス－その代表性と類似性」『英語コーパス研究』 18, 49-59.
- Ullmann, Stephen. (1962) *Semantics: An introduction to the science of meaning*. Oxford: Basil Blackwell.

付録：コーパス検索結果

※表中の表記：SECTION = 年代, FREQ = 頻度, WORDS(M) = 各年代の総語数 (単位：100 万), PER MIL = 100 万語あたりの調整頻度, 最下段 = PER MIL の棒グラフ

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	37178	17	68	89	121	351	288	381	535	824	1460	2658	2793	3138	3138	2949	2748	3628	3602	4053	4337
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	91.80	2.44	4.96	5.63	7.32	20.72	15.33	18.99	26.19	37.49	63.19	103.42	100.80	114.53	109.49	101.26	95.32	121.53	108.66	116.39	122.33
SEE ALL YEARS AT ONCE																					

COHA における international の年代ごとの分布

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	535	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	88	272	168
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	1.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	2.65	7.81	4.74
SEE ALL YEARS AT ONCE																					

COHA における globalization の年代ごとの分布

SECTION	ALL	1820	1830	1840	1850	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	16658	0	0	0	0	0	0	3	1	0	2	1	1	4	9	98	649	1475	3176	4472	3664	3103
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	41.13	0.00	0.00	0.00	0.00	0.00	0.16	0.05	0.05	0.00	0.09	0.04	0.04	0.14	0.33	3.42	22.29	51.16	106.39	134.90	105.22	87.52
SEE ALL YEARS AT ONCE																						

COHA における computer の年代ごとの分布

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	6074	0	0	1	6	0	2	2	5	2	2	0	5	2	1	0	3	2	1327	2683	2032
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	15.00	0.00	0.00	0.06	0.36	0.00	0.11	0.10	0.24	0.09	0.00	0.19	0.07	0.04	0.03	0.00	0.10	0.07	40.03	77.05	57.32
SEE ALL YEARS AT ONCE																					

COHA における internet の年代ごとの分布

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	6198	0	0	0	0	0	0	0	0	0	0	0	1	2	6	1	4	15	705	1997	3467
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	15.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.21	0.03	0.14	0.50	21.27	57.35	97.79
SEE ALL YEARS AT ONCE																					

COHA における online の年代ごとの分布

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	810	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	211	564
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.06	6.06	15.91
SEE ALL YEARS AT ONCE																					

COHA における blog の年代ごとの分布

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	5824	12	38	45	139	153	166	154	183	198	341	425	261	369	247	264	309	424	669	744	683
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	14.38	1.72	2.77	2.85	8.41	9.03	8.84	7.67	8.96	9.01	14.76	16.54	9.42	13.47	8.62	9.07	10.72	14.20	20.18	21.37	19.27
SEE ALL YEARS AT ONCE																					

COHA における mouse の年代ごとの分布

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	7302	37	86	103	78	154	149	122	103	124	117	123	132	122	155	142	167	253	1222	2518	1395
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	18.03	5.30	6.27	6.52	4.72	9.09	7.93	6.08	5.04	5.64	5.06	4.79	4.76	4.45	5.41	4.88	5.79	8.48	36.86	72.31	39.35
SEE ALL YEARS AT ONCE																					

COHA における web の年代ごとの分布

SECTION	ALL	1790s	1800s	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
FREQ	979	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	650	242
WORDS (M)	129	0.6	0.4	1.0	1.6	2.6	2.3	3.3	2.9	6.0	7.8	10.1	7.3	5.9	4.5	4.8	6.7	4.7	7.7	12.0	14.4	9.2	7.5	6.1
PER MIL	7.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.49	86.18	39.42
SEE ALL SUB- SECTIONS AT ONCE																								

Corpus of US Supreme Court Opinions における internet の年代ごとの分布

SECTION	ALL	1790s	1800s	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
FREQ	572	0	0	1	1	0	0	1	0	2	60	11	6	0	5	32	21	8	17	17	26	44	217	103
WORDS (M)	129	0.6	0.4	1.0	1.6	2.6	2.3	3.3	2.9	6.0	7.8	10.1	7.3	5.9	4.5	4.8	6.7	4.7	7.7	12.0	14.4	9.2	7.5	6.1
PER MIL	4.42	0.00	0.00	0.97	0.62	0.00	0.00	0.30	0.00	0.33	7.70	1.09	0.83	0.00	1.10	6.66	3.15	1.71	2.22	1.42	1.81	4.80	28.77	16.78
SEE ALL SUB- SECTIONS AT ONCE																								

Corpus of US Supreme Court Opinions における web の年代ごとの分布

